Accounting for model error due to unresolved scales within ensemble Kalman filtering

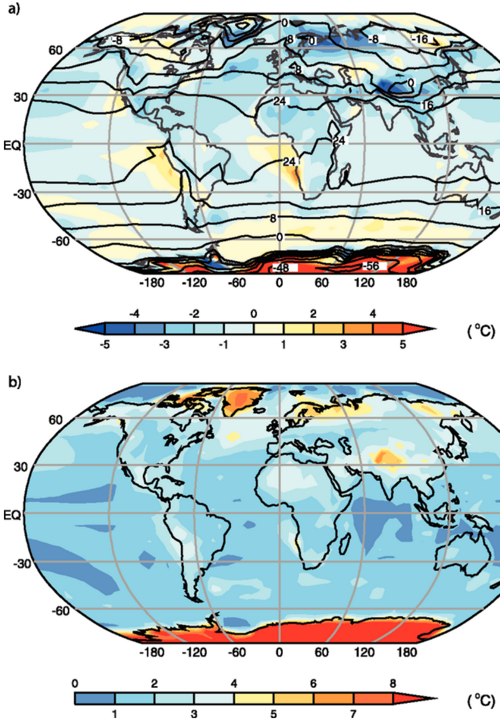
Lewis Mitchell & Alberto Carrassi (IC3/NERSC)

THE UNIVERSITY of ADELAIDE

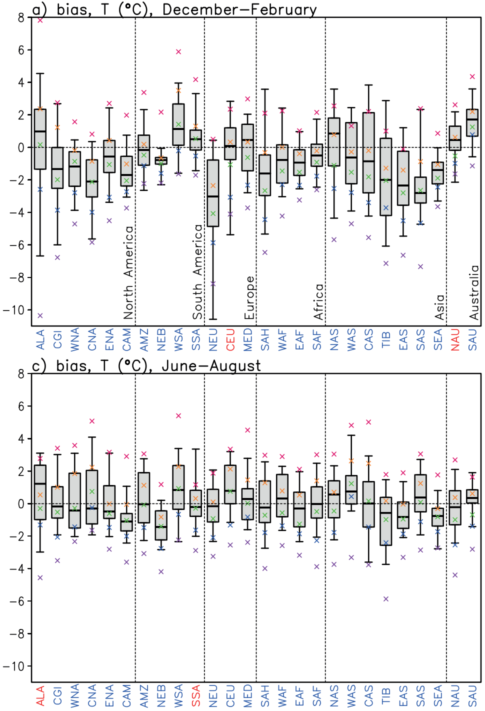
SUB CRUCE LUMEN

maths.adelaide.edu.au/lewis.mitchell

@lewis_math



(IPCC AR4 Fig. 8.2)



(IPCC AR5 Fig. 9.39)

Outline

Idea

Comparison with data shows that all models are wrong.
Can we use data assimilation to incorporate model error/improve models?

1. Deterministic approach to model error
2. Data assimilation and 2 new model error strategies
3. Numerical results

Problem formulation (Nicolis 2004)

Model:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, \boldsymbol{\lambda})$$

Truth:

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = \hat{\mathbf{f}}(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}}) \qquad \frac{d\hat{\mathbf{y}}(t)}{dt} = \hat{\mathbf{g}}(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}})$$

$$\mathbf{x}, \hat{\mathbf{x}} \in \mathbb{R}^N,\ \hat{\mathbf{y}} \in \mathbb{R}^M$$

Formal (unrealisable) solutions:

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t d\tau \mathbf{f}(\mathbf{x}(\tau), \boldsymbol{\lambda}) \tag{1}$$

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}_0 + \int_0^t d\tau \hat{\mathbf{f}}(\hat{\mathbf{x}}(\tau), \hat{\mathbf{y}}(\tau), \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}}) \tag{2}$$

Problem formulation (Nicolis 2004)

Averaging $\mathbf{x} - \hat{\mathbf{x}}$ over ensemble of initial conditions:

$$\langle \delta\mathbf{x}(t) \rangle = \langle \delta\mathbf{x}_0 \rangle + \int_0^t d\tau \left\langle \mathbf{f}(\mathbf{x}(\tau), \boldsymbol{\lambda}) - \hat{\mathbf{f}}(\hat{\mathbf{x}}(\tau), \hat{\mathbf{y}}(\tau), \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}}) \right\rangle \quad (3)$$

Assuming unbiased initial conditions, short-time interval:

$$\mathbf{b}_m = \langle \delta\mathbf{x}(t) \rangle \approx \left\langle \mathbf{f}(\mathbf{x}(\tau), \boldsymbol{\lambda}) - \hat{\mathbf{f}}(\hat{\mathbf{x}}(\tau), \hat{\mathbf{y}}(\tau), \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}}) \right\rangle t$$

$$\mathbf{P}_m(t) \approx \left\langle \left\{ \mathbf{f}(\mathbf{x}_0, \boldsymbol{\lambda}) - \hat{\mathbf{f}}(\hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}}) \right\} \left\{ \mathbf{f}(\mathbf{x}_0, \boldsymbol{\lambda}) - \hat{\mathbf{f}}(\hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\epsilon}}) \right\}^T \right\rangle t^2$$

Model error strategies (Carrassi & Vannitsem 2011)

Assume we have access to a high-quality reanalysis time series $\mathbf{x}^a(t)$:

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}, \lambda) - \hat{\mathbf{f}}(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \lambda, \hat{\epsilon}) &= \frac{d\mathbf{x}}{dt} - \frac{d\hat{\mathbf{x}}}{dt} \\
&\approx \frac{\mathbf{x}_r^f(t+\tau_r) - \mathbf{x}_r^a(t)}{\tau_r} - \frac{\mathbf{x}_r^a(t+\tau_r) - \mathbf{x}_r^a(t)}{\tau_r} \\
&= \frac{\mathbf{x}_r^f(t+\tau_r) - \mathbf{x}_r^a(t+\tau_r)}{\tau_r} \\
&= -\frac{\delta\mathbf{x}_r^a}{\tau_r}
\end{aligned}
$$

where $\delta\mathbf{x}_r^a$ is a *reanalysis increment*. Then

$$
\overline{\mathbf{b}}_m = \langle \delta\mathbf{x}_r^a \rangle \frac{\tau}{\tau_r}
$$

$$
\overline{\mathbf{P}}_m = \left\langle \{\delta\mathbf{x}_r^a - \langle\delta\mathbf{x}_r^a\rangle\}\{\delta\mathbf{x}_r^a - \langle\delta\mathbf{x}_r^a\rangle\}^T \right\rangle \frac{\tau^2}{\tau_r^2}
$$

Model error strategies: implementation

Time-constant method: **ETKF-TC**

$$\mathbf{x}^f \Longrightarrow \mathbf{x}^f - \alpha\bar{\mathbf{b}}_m,$$

$$\mathbf{P}^f \Longrightarrow (1+\delta)\mathbf{P}^f \circ \mathbf{\Omega}(r) + \alpha^2\overline{\mathbf{P}}_m$$

δ: covariance inflation, $\mathbf{\Omega}$: localization, α: model error 'inflation'

Time-varying method: **ETKF-TV** (LM & Carrassi 2014)

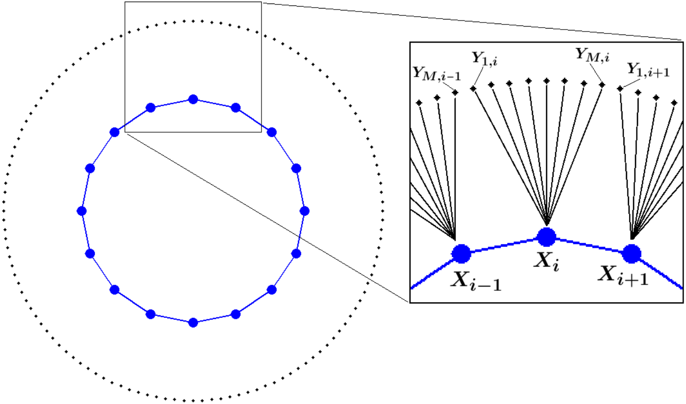
$$\mathbf{x}^f_{i,j} = \mathcal{M}(\mathbf{x}^a_{i,j}) - \alpha\boldsymbol{\eta}_{i,j}\frac{\tau}{\tau^r} \qquad \boldsymbol{\eta}_{i,j} \in \mathcal{N}(\overline{\mathbf{b}}_m, \overline{\mathbf{P}}_m) \qquad i = 1, ..., k$$

cf. stochastic climate models (Harlim & Majda 2010, LM & Gottwald 2012)

Lorenz-96 model

$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F - \frac{hc}{b}\sum_{j=1}^{J} y_{i,j}$$

$$\dot{y}_{i,j} = -cb(y_{i,j+2} - y_{i,j-1})y_{i,j+1} - cy_{i,j} + \frac{hc}{b}x_i$$

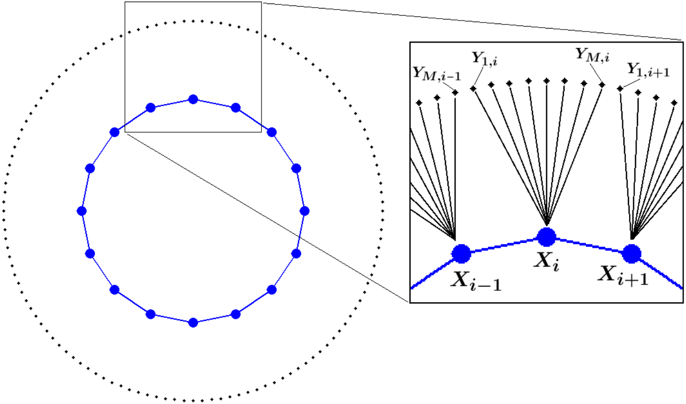


(From Bellsky, Berwald, & LM 2014)

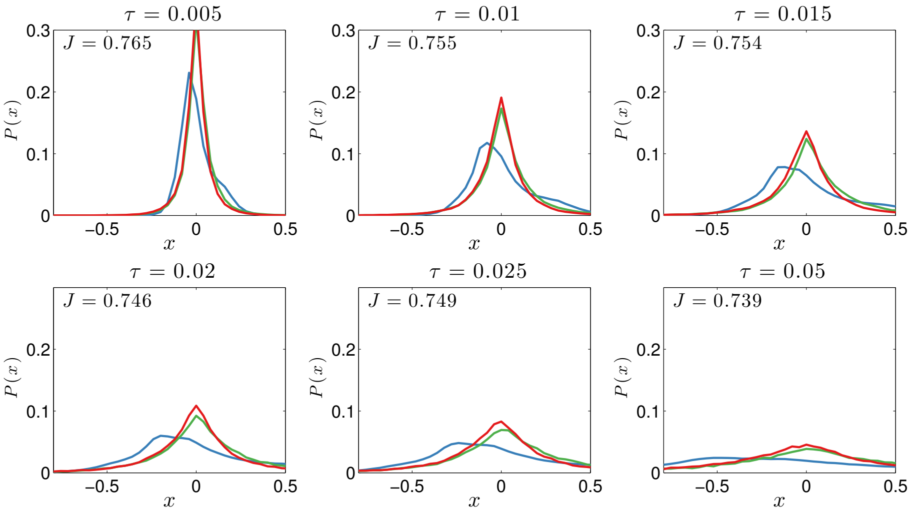
Lorenz-96 model

$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F - \frac{hc}{b}\sum_{j=1}^{J} y_{i,j}$$

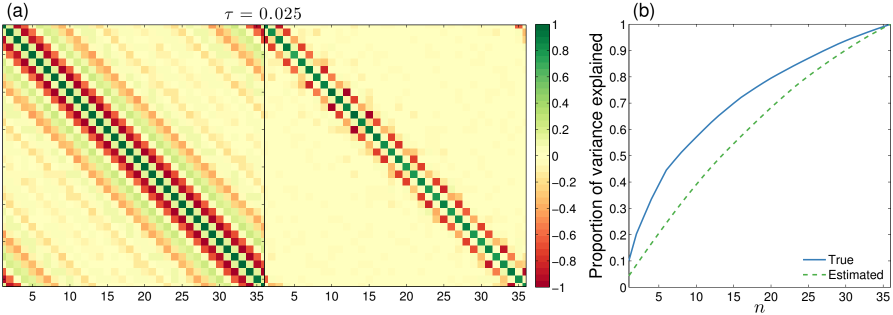
$$\dot{y}_{i,j} = -cb(y_{i,j+2} - y_{i,j-1})y_{i,j+1} - cy_{i,j} + \frac{hc}{b}x_i$$



(From Bellsky, Berwald, & LM 2014)



red: observed, green: unobserved, blue: true



(a)
$\tau = 0.025$
5
10
15
20
25
30
35
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
(b)
Proportion of variance explained
n
True
Estimated

Experimental setup

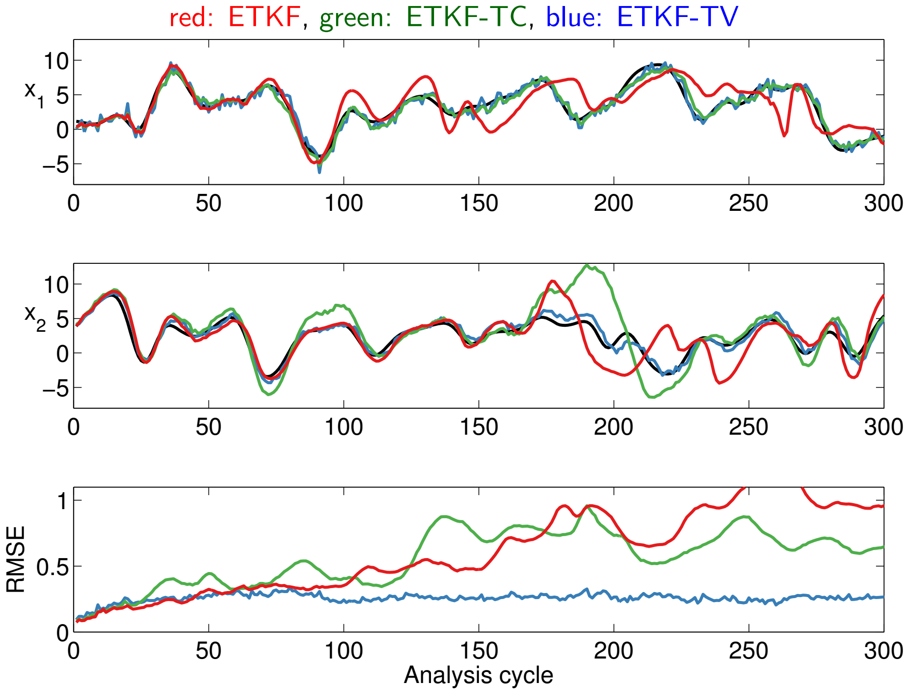
Reanalysis phase:

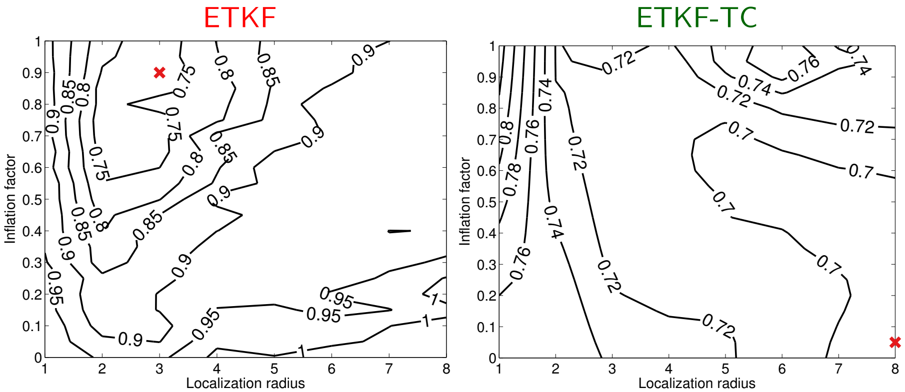
- 10-year reanalysis dataset (forecast + analysis means) using optimised (best inflation δ and localization radius r)
- observations every $\tau = 3$hr, every 3rd variable
- obs. noise 5% climatological variance

Experiment phase:

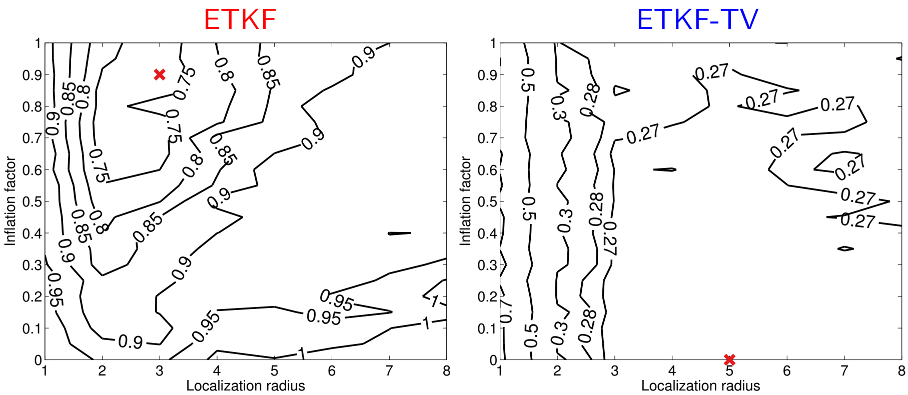
- Same parameters as reanalysis, 1 month experiment
- Average results over different initial background ensemble only
- ETKF-TC/TV use $\overline{\mathbf{P}}_m$ generated by optimized ETKF

Same truth + observations for all experiments (one Earth!)


red: ETKF, green: ETKF-TC, blue: ETKF-TV
x_1
x_2
RMSE
Analysis cycle
0
50
100
150
200
250
300
10
5
0
−5
1
0.5

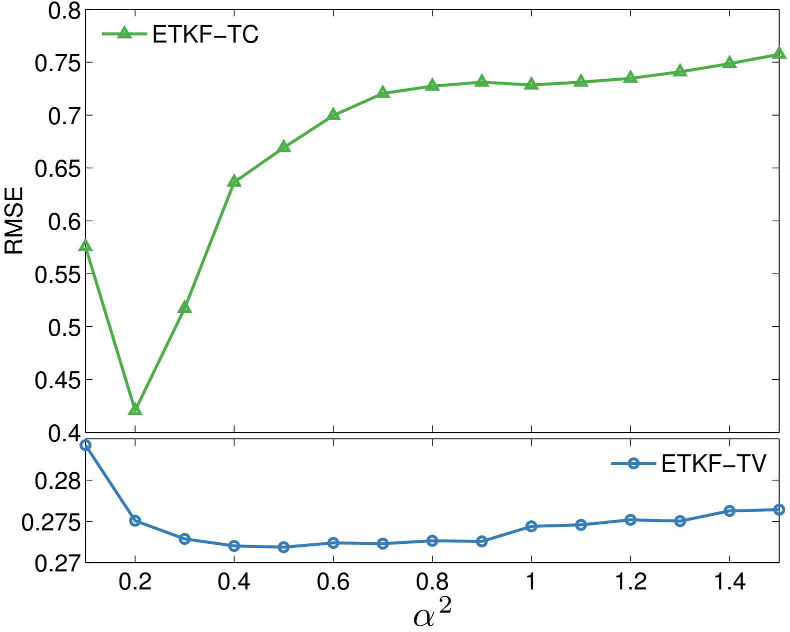


ETKF-TC comparable to optimally tuned ETKF,
outperforms sub-optimally tuned ETKF

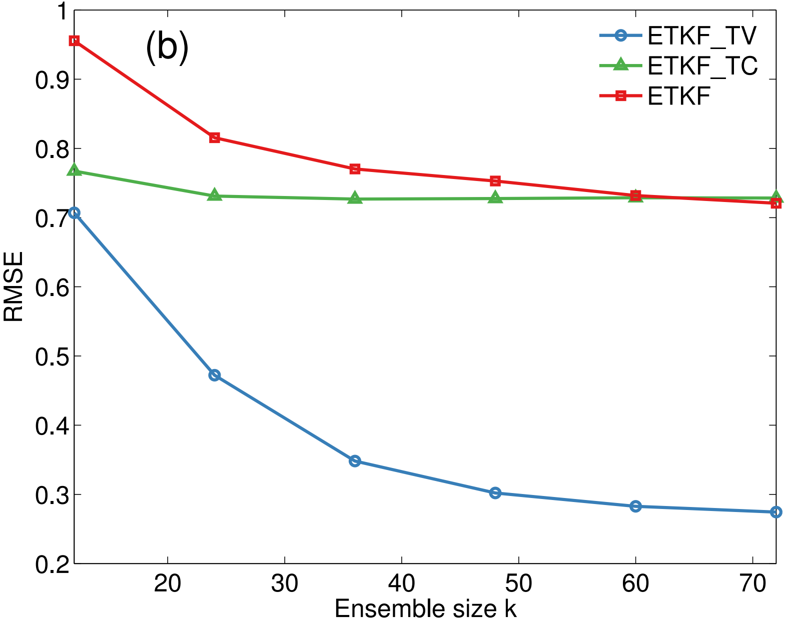


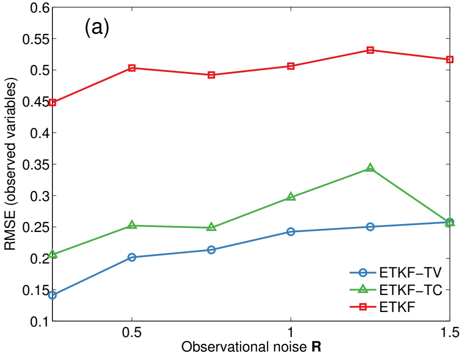
ETKF-TV always outperforms ETKF,
insensitive to inflation/localisation tuning

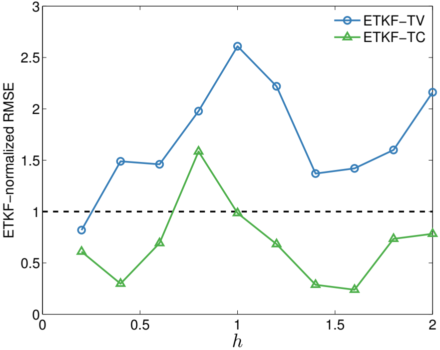
...but you can tune ETKF-TC to account for initial condition error:



...and it works well for small ensemble sizes (which is important!):




(a)
RMSE (observed variables)
Observational noise **R**
ETKF−TV
ETKF−TC
ETKF
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.5
1
1.5


ETKF−normalized RMSE
h
ETKF−TV
ETKF−TC
0
0.5
1
1.5
2
2.5
3
0
0.5
1
1.5
2

Summary & Outlook

We have:

- Described a deterministic, short-time approach to the model error problem
- Approximated correction terms using reanalysis data
- Proposed 2 implementations for ETKF, which
  - are robust to parameter tuning
  - perform well for small ensemble sizes

We will:

- Apply ETKF-TC/TV up the hierarchy of models
- Look at the problem of parameter estimation + model error
- Disentangle model from initial condition error
- … more?